The Scale Argument for Longtermism

Tomi Francis

February 20, 2023

1 Introduction

Currently, there are around eight billion humans on earth. The number of humans who have ever lived stands at around one hundred billion.[1] That's a lot of people, but it's peanuts compared to the number of people who might exist in the future. Greaves and MacAskill (2021: 9) "conservatively" suggest that there are at least 10^{24} future people in expectation. Suppose they're right about that. Does this matter for our moral decision making in the here and now? Greaves and MacAskill think that it does. They argue that *because* the expected number of future lives is so incomprehensibly vast, we should accept Longtermism: roughly, the view that the beneficial effects we can have on the far future are vastly more morally important than beneficial effects we can have in the here and now.[2] This basic case for Longtermism has been summarised pithily by MacAskill (2022):

> Future people count. There could be a lot of them. We can make their lives go better.[3]
>
> MacAskill (2022: 9)

As arguments go, this is light on detail. But it does still help categorise some of the ways opponents of Longtermism might resist the conclusion. When it comes to the first premise, the main objections are theoretical. For instance, one might defend a pure rate of time preference, on which people in the very far future count so little they might as well be ignored. One might argue that paradigmatic Longtermist interventions like the mitigation of extinction risk

[1] Kaneda and Haub (2022).

[2] Strictly speaking, I'm talking about the view Greaves and MacAskill call "Strong Longtermism". I'll stick with "Longtermism" for this view since that's less clunky, but note that this view is stronger and more controversial than "Longtermism" as advocated by MacAskill (2022), which holds merely that "positively influencing the far future is *a* key moral priority of our time" [emphasis mine].

[3] This last claim should raise the eyebrows of anyone familiar with the Non-Identity Problem. Because our actions change *who* will exist in the future, it is false that we can make the lives of future people go better, understood in the *de re* sense. However, we plausibly can make future people's lives go better in the *de dicto* sense, and those philosophers who accept Longtermism think that we have very strong moral reasons to do so.

have the effect of increasing the expected number of happy people, and that *this* kind of change isn't a change for the better.[4] Or, one might argue that we cannot provide future people with comparative benefits due to the non-identity problem, and that we can only have a moral reason to provide comparative benefits.[5]

In contrast, the second and third premises might be rejected on empirical grounds. Projections of 10^{24} future lives, or the continuation of humanity for millions of years, might be attacked as being unfounded science fiction-esque speculation. Or, it might be argued that the background rate of extinction is either too small for us to make much difference to it, or too large to result in a massive expected future population.[6]

Finally, they might deny the third premise, that we can make future people's lives better.[7] They might say that we've just got no idea how our actions will affect the far future, so changing it for the better isn't something we can realistically aim at.[8]

Setting these objections aside, we can ask whether the argument succeeds. Call it the Scale Argument for Longtermism: *because* there could be a vast number of future people, the effects of our actions on the far future (and therefore on this vast number of future people) are overwhelmingly morally important. Versions of the Scale Argument have been given before. Usually, they begin with a utilitarian calculation: one simply tots up the expected aggregate wellbeing occurring in the future, and compares it to the expected aggregate wellbeing occurring in the past. On a calculation like this, the far future can't help but win out once the numbers get large enough: more people means more wellbeing at stake.

It might at first appear that this argument should only convince committed utilitarians: a mere fraction of professional moral philosophers, and probably a smaller faction of people on the proverbial street. But this appearance is deceptive. As Greaves and MacAskill (2021) point out, the *very large* difference in expected value at stake, and the variety of ways in which we might make the future better, makes the Longtermist conclusion robust to many plausible variations in moral theory. You're a Prioritarian? That's fine! The far future also contains much more priority-weighted utility. You accept the Procreation Asymmetry? No problem! We can make the future better by focusing on making the future better if we *don't* go extinct, thereby increasing the average quality of life rather than increasing the number of happy people. You're risk averse? That's ok! Since Longtermist interventions usually aim at mitigating risks,

[4] See Finneron-Burns (2017), Frick (2017) and Frick and Lederman (forthcoming).

[5] Boonin (2014) offers a defence of a view like this.

[6] See Thorstad (2022) for a detailed version of this argument.

[7] Assuming existence non-comparativism (the view that an outcome in which a person exists cannot be better or worse for her than an outcome in which she does not exist), given the non-identity problem, we clearly cannot make future people's lives better in the *de re* sense. To be charitable, then, this premise should be read in the *de dicto* sense: we can make future people's lives better in the sense that we can cause better-off rather than worse-off people to exist.

[8] See Lenman (2000) and Greaves (2016).

it might actually strengthen the case; if not, the vast difference in aggregate wellbeing will override any risk-aversion pulling in the opposite direction. And so on.

Still, there is a deeper worry one might have with the sorts of utilitarian calculations presented by Bostrom (2003) and Greaves and MacAskill (2021). One might think that doing utilitarian calculations is just the wrong way to go about doing ethics. If your view isn't even in the same ballpark as utilitarianism to begin with, a utilitarian calculation is unlikely to convince you.

The aim of this paper is to rectify this deficiency in previous versions of the Scale Argument for Longtermism. I'm going to present a version of the Scale Argument which doesn't involve adding up people's wellbeing in two alternatives and checking which is the bigger number. Instead, I'm going to argue that we can get to Longtermism by iteratively applying uncontroversial principles to the effect that effects on future people matter *at least a little bit*. By applying these principles many times, we're going to find that if the future is large enough, it must matter *a lot*. For the purposes of this paper, I won't take much notice of the empirical objections to Longtermism. The point is rather to see whether it's true that *if* Longtermists are right about empirical matters, like the future being "large enough", then Longtermism indeed follows on any plausible moral theory.

Following Greaves and MacAskill (2021), I will divide my discussion into two parts. In the first part, I will talk about axiological Longtermism: would it be *better* to improve the future than to improve the present? In the second part, I shall talk about deontic Longtermism: do we have a moral obligation to improve the future rather than the present, given a binary choice between the two? The difference between the two discussions mainly comes down to transitivity. When it comes to axiology, it is widely (though not universally) accepted that if A is at least as good as B, and B is at least as good as C, A must be at least as good as C. This principle of transitivity allows my iterative argument to get going. In contrast, the analogous deontic principle would state that if A ought to be chosen rather than B, and B ought to be chosen rather than C, then A ought to be chosen rather than C. This principle does have something to be said for it, but it's much more controversial than its axiological counterpart. My discussion in [SECTION] will mostly focus on how the iterative argument can be adapted for the non-transitive case, using two nice tricks developed by Arrhenius (2022) and Hare (2016).

2 Three Types of Existential Risk Reduction

Before we begin, it's worth clearing up a common misconception about the case for Longtermism. This is that the only way we could possibly affect the far future is to reduce the risk of human extinction; consequently, the Scale Argument for Longtermism must involve the assumption that it would be a good thing to increase the number of happy people. But that's a controversial assumption which has been rejected by many philosophers, especially in the

context of the importance of averting human extinction.[9]

Longtermists generally claim that, in addition to affecting the far future by reducing extinction risk, we can also affect the future by increasing the average quality of life. For example, Greaves and MacAskill (2021) argue that we have at least two paths to long-term impacts of this sort. The first is work to reduce the risk that the development of Artificial General Intelligence locks us in to a situation in which poor values predominate (for example, if this technology is used to maintain and empower a fascist dictatorship). The second is to invest money for the long term, using the power of compounding returns to transform relatively small sums of money today into comparatively vast resources which future altruists might use to tackle the problems of their time.

For my part, it seems to me plausible that there will be at least something in the vicinity we can do which might avert a disaster not involving human extinction with some small, but appreciable, probability. But I won't argue for this conclusion. Again, our main question is: *if* Longtermists are right about their empirical claims, does Longtermism follow on any reasonable moral theory?

Let us then distinguish three sorts of changes we might make to the future. An *extinction-mitigating* change reduces the risk of human extinction. A *quality-improving* change has some chance of producing a future in which better-off people exist, rather than a future in which roughly the same number of worse-off (non-identical) people exist, given that in both cases the future people have lives worth living. Finally, a *suffering-reducing* change reduces the chance that a horrible future involving large numbers of lives not worth living will come about, replacing this with a future in which humans still exist, but have lives worth living on the whole.

I have listed these three sorts of changes in increasing order of the plausibility that making such a change (or rather, the effects of these changes on people) would be morally important. It is controversial whether extinction-mitigating changes matter, in so far as they affect people.[10] It is somewhat less controversial that quality-improving changes matter: few people would regard it as a matter of indifference if the whole of humanity were to be infected by a virus whose only effect was to reduce the lifespans of our great-great-great-great grandchildren by half, even assuming that all significant sources of suffering will be reduced by that point in time. Still, due to the Non-Identity Problem, some philosophers disagree on this issue.[11] Suffering-reducing changes, however, enjoy almost unanimous support. Practically everyone (rightly) agrees that it would be bad for people to come into existence with lives full of uncompensated suffering, and then die.

[9] Sceptics of the moral importance of creating happy people include Narveson (1967), Frick (2017), Finneron-Burns (2017) and Roberts (2011). Proponents include Chappell (2017) and (in the axiological case) Broome (2004, 2005).

[10] Frick (2017) gives an account of why preventing extinction might be morally important even if the associated effect on people (i.e., the creation of people who exist after the averted extinction event) is not.

[11] See Boonin (2014).

I shall therefore focus on suffering-reducing changes. What I shall say can, going forward, be adapted to the other two cases with minimal effort. Why bother considering the other two types of change? Well, you might think that the empirical case for our being *able* to make the other two kinds of changes is more robust. This is especially the case for extinction-mitigating changes: by considering the version of my argument involving changes of this sort, it seems that one would make the case for Longtermism more empirically robust, at the cost of making it substantially less robust to differing moral theories.

3 The Axiological Case

3.1 Setup

Let us then turn to the case for axiological Longtermism. Here's the setup. There are around eight billion people on earth now. To make things a bit neater, let's round that up and imagine that there are ten billion. Assume that if nothing is done, each of these present people will get a life which is mediocre throughout, just barely above the neutral level. (Call this wellbeing level ϵ.)

Humanity might go prematurely extinct, in which case there won't be many future people. But I shall assume that there is some probability p, which is at least one in ten thousand, that humanity will not go prematurely extinct. In that case, when it comes to future people, I shall assume that

> *There Are A Lot Of Them* There will, with probability p, be at least 10^{24} future people.[12]

Finally, imagine that that if nothing is done, there is a $\frac{1}{10,000}$ probability that all 10^{24} future people will have lives at wellbeing level -100 (corresponding to 100 years of life in atrocious conditions). Now consider

> *Short-Term vs Long-Term Intervention* A moral agent has one hundred billion dollars to spend on either a *short-term intervention* S or a *long-term intervention* L. No mixed strategies are possible.
>
> S will help present people, shifting them from ϵ to 100 units of wellbeing (corresponding to 100 years of high quality life). However, S will do nothing to improve the future.
>
> L will do the opposite: it will do nothing to help present people, but will avert the $\frac{1}{10,000}$ chance of us having a terrible future. If L is done, then instead of having 10^{24} future people at level -100, we will instead have 10^{24} different future people at level ϵ.

[12] For comparison, Newberry (2021) estimates a future population of 10^{27} for the case in which humanity spreads throughout the solar system, but not beyond, and in which humanity remains biological. The numbers get much larger if one relaxes the assumption that humanity is *forever* bound to the Solar System.

Short-Term vs Long-Term Intervention is much simplified compared to actual choice situations any person or institution faces today. But it is simplified in ways which mostly favour the case for choosing the short-term intervention. S involves giving eight billion people an excellent life, in place of one that is barely worth living. Presumably, this is roughly as good as saving eight billion lives. That's better than anything we can actually do to affect the short term. For one thing, most humans don't need their lives saving. For another, even supposing the best short-term interventions could somehow be scaled up this far without a massive reduction in cost-effectiveness (which is far from true in practice), the cost would be prohibitive. GiveWell estimates that one of the most cost-effective charities aimed at saving lives today, the Against Malaria Foundation, saves roughly one life per $4,000 in Guinea.[13] Scaling this up to eight billion, the cost of an S-like intervention would be *at least* twenty four trillion dollars. The cost-effectiveness of the short-term intervention is therefore dramatically overstated in our example.

In contrast, the cost-effectiveness of the long-term intervention is likely understated relative to the empirical claims generally made by longtermists. Ord (2020: 167) puts the total next-century risk at around one in six. The probability of a catastrophe that leads to future people persistently having terrible lives is presumably significantly lower than the risk of outright extinction. Even so, on the *assumption* that Longtermists are broadly correct about the risks being of any appreciable magnitude, and about our ability to mitigate them, an effective deployment of one hundred billion dollars could presumably reduce the risk of this sort of catastrophe by more than one in a hundred thousand.

Overall, then, the simplifications involved in *Short-Term vs Long-Term Intervention* mostly favour choosing the short-term intervention. This justifies my choice of name for the following claim:

> *Axiological Longtermism* The prospect resulting from doing L in *Short-term vs Long-term Intervention* is better overall than the prospect resulting from doing S.

3.2 The Argument

I shall now present a version of the Scale Argument for Axiological Longtermism. It proceeds via two premises. The first premise is

> *Transitivity* If prospect X is better than prospect Y, and Y is better than prospect Z, then X is better than Z.

Additionally, we will need to assume that a small chance of one present person getting an extra year of good life is less important than averting the same small chance of one hundred million future people having atrocious lives. More precisely, the Scale Argument will appeal to

[13] GiveWell (2022). Cost-effectiveness is lower in the other countries in which AMF operates.

Future People Count Suppose that prospects X and Y differ in only the following ways:

(i) For some state of nature s_i with probability $\frac{1}{10{,}000}$, one present person has wellbeing level at most $w+1$ in X, rather than having wellbeing level w in Y, where $w \in [0, 99]$.

(ii) In some state of nature s_j (which may not be equal to s_i) with probability $\frac{1}{10{,}000}$, at least one hundred million future people exist with wellbeing level -100 in prospect X. In prospect Y, these hundred million future people do not exist. Instead, one hundred million different future people exist with wellbeing level ϵ.

Then Y is better overall than X.

We'll come back to the question of whether we *should* accept these two premises later, but let me first say something in their favour. Transitivity is a standard assumption about the betterness relation. Some philosophers dispute it, but they are far in the minority. Let's just take it for granted for the time being. As for *Future People Count*, let me just point out that this is a very weak claim. It does not require that future people matter anything like as much as present people It just requires that they matter a little bit, evaluatively speaking: enough that a $\frac{1}{10{,}000}$ chance of one hundred million of them living suffering-filled lives is more important than the same chance of a single present person losing one year of good life.

Let's now see how we can get from these two premises to Axiological Longtermism. Very roughly, the idea is this. We're going to use *Future People Count* to trade off small risks of small changes to the wellbeing of individual present people against same-sized risks of large identity-affecting changes to the wellbeing of future people. *Future People Count* will tell us that the risks to future people matter more when the stakes are low for present people in this way. However, by iterating these changes, we will eventually find ourselves trading off certainty of large changes in wellbeing of *all* present people against small risks of large identity-affecting changes to the wellbeing of a sufficiently large number of future people. Transitivity will then tell us that, since each small change is for the better, the combination of changes is also for the better. In particular, we will find that the prospect resulting from choosing L is better than the prospect resulting from choosing S.

Now for the details. We begin with a prospect X_0 (or X for short), in which ten billion present people get 100 units of wellbeing for sure and there is a $\frac{1}{10{,}000}$ chance of 10^{24} future people having atrocious lives. This is the prospect which results from the choice of the short-term intervention S. We then compare X to the prospect X_1, in which, with probability $\frac{1}{10{,}000}$, one present person has wellbeing level 99 instead of 100; in return, one hundred million (10^8) of the possible people who would have atrocious lives in the future are replaced by the same number of different future people with mediocre lives at ϵ. According to *Future People Count*, X_1 is better overall than X.

We then repeat this process ten thousand times, each time giving the one present person a slightly higher chance of getting wellbeing level 99 rather than 100. Prospect $X_{10,000}$ is the prospect in which the one present person is sure to get wellbeing level 99 rather than 100; in return, 10^{12} possible future people with atrocious lives are replaced by the same number of different possible future people with mediocre lives. We then repeat the entire process, moving this same person down to wellbeing level 98 (by the time we get to prospect $X_{20,000}$, then to 97, and so on until we get to $\epsilon < 1$ at $X_{1,000,000}$. In this prospect, one present person certainly has wellbeing level ϵ rather than 100; in return, 10^{14} possible future people with atrocious lives are replaced by the same number of different possible future people with mediocre lives. Finally, we repeat this entire process for each person in the present population of ten billion. The result, at prospect $X_{10,000,000,000,000,000}$, which we may call Y for short, has all present people at wellbeing level ϵ rather than 100, but averts the $\frac{1}{10,000}$ chance of 10^{24} future people having atrocious lives, replacing these people by different possible future people who would, if they exist, have lives at level ϵ. That is, Y is the prospect resulting from choosing the long-term intervention L. *Future People Count* implies that each $X_{i+1} \succ X_i$; transitivity thus implies that $Y \succ X$. Since Y is the prospect resulting from doing L and X is the prospect resulting from doing S, we conclude that Axiological Longtermism is true.

3.3 Discussion

How might one object to this argument? On the empirical side, maybe the future is not so large in expectation, or perhaps there just isn't a non-negligible risk of a terrible future coming to pass, or maybe we can't do anything about that risk even if it is non-negligible. These are all real concerns, but as I've said, I'll set them aside: we want to see whether Longtermism is plausible *if* Longtermists are broadly correct about their empirical claims.

There are two non-empirical claims involved in the version of the Scale Argument for Longtermism I've just presented. The first is transitivity: if X is better than Y, and Y is better than Z, then X is better than Z. The second is the claim that *Future People Count*.

As I've pointed out, *Future People Count* is quite a weak claim: it only requires that future people matter *a little bit*. There are still several views on which *Future People Count* is false. I shall briefly outline them, then explain why I think they should be rejected.

[I shall discuss pure time discounting, axiological partial aggregation (Lazar and Lee-Stronach (2019) EUT-compliant version), narrow person-affecting views and radical different-number incomparability. My line will mostly be, "these theories just aren't plausible *because* they either say we shouldn't care about the future at all, or they say that the extent to which we should care about future people depends in implausible ways on how many people we've already helped or exactly *how far* in the future they are.]

[In relation to rejecting transitivity, I shall discuss essentially comparative value and axiological partial aggregation (cyclic version). I shall point out that

most people reject non-transitive value, and apart from that I shall say that what I shall say in the section on the Deontic case applies, changing what needs to be changed, to the case of non-transitive axiology.]

4 The Deontic Case

We have seen an argument for Axiological Longtermism. Consider now its counterpart,

> *Deontic Longtermism* In Short-term vs Long-term Intervention, one ought to do L rather than S.

Can the Scale Argument for Axiological Longtermism be adapted for Deontic Longtermism? Here's one obvious thought. Suppose we simply replace the "better than" relation $\succ$ by a deontic relation $\succ_D$, which we might take to be the relation "[...] ought to be chosen rather than [...] if these are the only options". We can uniformly replace all instances of $\succ$ by the new relation $\succ_D$ in the Scale Argument for Axiological Longtermism. The result will of course be a valid argument for Deontic Longtermism which appeals to the following two premises:

> *Deontic Transitivity* For any prospects X, Y and Z, if $X \succ_D Y$ and $Y \succ_D Z$, then $X \succ_D Z$.
>
> *Future People Count (Deontic)* Suppose that prospects X and Y differ in only the following ways:
>
> (i) For some state of nature s_i with probability $\frac{1}{10{,}000}$, one present person has wellbeing level at most $w+1$ in X, rather than having wellbeing level w in Y, where $w \in [0, 99]$.
>
> (ii) In some state of nature s_j (which may not be equal to s_i) with probability $\frac{1}{10{,}000}$, at least one hundred million future people exist with wellbeing level -100 in prospect X. In prospect Y, these hundred million future people do not exist. Instead, one hundred million different future people exist with wellbeing level ϵ.
>
> Then $Y \succ_D X$.

The resulting argument will be valid because, just as before, Future People Count (Deontic) will imply that $X_{i+1} \succ_D X_i$ for each i, and Deontic Transitivity will allow us to deduce from these pairwise claims that $Y \succ_D X$. However, it is not an argument that will persuade most non-consequentialists, because Deontic Transitivity is widely disputed on independent grounds; see for example (CITATIONS). The deontic analogue of the Scale Argument for Axiological Longtermism is therefore not persuasive. This is the case because of a structural difference between value and obligation: betterness is transitive, whereas obligation in two-option choice sets may not be transitive.

If the Scale Argument for Longtermism is to be revived in the deontic case, it will have to get by without transitivity. And as it happens, philosophers have proposed quite general methods of taking iterative arguments, which might work straightforwardly in the context of a transitive axiology, and transforming them into arguments that work for deontic notions like obligation, without assuming that $\succ_D$ is in general a transitive relation.

4.1 The "No Moral Dilemmas" Approach

One method has recently been put to work in the context of the paradoxes of population ethics by Gustaf Arrhenius (2022). The idea is to reinterpret axiological claims of the form $Y \succ X$ as instead saying that, for *any* choice set in which X and Y are among the options, X is wrong. Reinterpreted this way, our axiological premise that *Future People Count* becomes

> *Future People Count (NMD version)* Suppose that prospects X and Y differ in only the following ways:
>
> (i) For some state of nature s_i with probability $\frac{1}{10{,}000}$, one present person has wellbeing level at most $w+1$ in X, rather than having wellbeing level w in Y, where $w \in [0, 99]$.
>
> (ii) In some state of nature s_j (which may not be equal to s_i) with probability $\frac{1}{10{,}000}$, at least one hundred million future people exist with wellbeing level -100 in prospect X. In prospect Y, these hundred million future people do not exist. Instead, one hundred million different future people exist with wellbeing level ϵ.
>
> Then, in any choice set in which X and Y are among the options, X is wrong.

I think that the NMD version of Future People Count is about as plausible as its axiological counterpart. It seems clearly important to prevent a one in ten thousand risk of *one hundred million* future people from suffering. It certainly seems more important than preventing the same small risk of a single present person being made slightly worse off. And, at first glance, the presence of other options doesn't seem like it should alter this judgement.

By applying Future People Count (NMD version) repeatedly to the prospect pairs X_i, X_{i+1}, we can conclude that in the option set $\{X_0, X_1, \cdots, X_{10^{16}}\}$, all options except (perhaps) $X_{10^{16}}$ are wrong. This brings us to our next step, which is not to apply any version of transitivity. Instead, we will apply

> *No Moral Dilemmas* In any choice set, at least one option is permissible.

Since $X_{10^{16}}$ is the only option that *might* be permissible, No Moral Dilemmas implies that it *is* permissible, and hence obligatory.

You might be thinking that this suffices to show that L is right and S is wrong in *Short-Term vs Long-Term Intervention*. However, that would be too

quick. It *would* be enough to show that L is obligatory if the intermediate options corresponding to prospects X_1, X_2 and so on were available. But they are not. And these intermediate prospects are not in any realistic extension of the case, either. By allocating varying amounts of money between short-term and long-term interventions, we might be able to achieve lesser or greater reductions in various existential risks. But it's doubtful that we would be able to prevent a small risk of *part* of the future population from suffering with our limited resources, when we might instead prevent the same risk of a larger part of the future population from suffering with a greater allocation of resources.

To get to the conclusion that L is obligatory, then, we need another principle, namely

> *Obligation-Contraction Consistency* If X is obligatory in some choice set C, then X is obligatory in any smaller choice set $C' \subset C$.

This principle directly implies that, since $X_{10^{16}}$ is obligatory in the larger option set in which all intermediate prospects are available, it remains obligatory in the smaller option set $\{X_0, X_{10^{16}}\}$; hence L is obligatory in *Short-Term vs Long-Term Intervention*. The thought behind this principle is that if we have determined what we ought to do on a generous conception of the options available, we don't then need to go and work out whether all of these options are genuinely feasible in order to check that our determination was correct. Intuitively, it doesn't matter whether potential impermissible options are actually available.

So, we have ourselves a valid argument for Deontic Longtermism. But is it a good argument? Well, the NMD version of *Future People Count* certainly looks plausible. But perhaps one might object on the basis that, as the chances of small harms to present people add up to certainty of large harms, it becomes less obvious that small chances of additional minor harms to present people can be brushed off as insignificant compared to the apparently much greater stakes for future people.

Perhaps an analogy might help make this point clear. It is obvious that we should prevent ten thousand people from suffering a minor headache, rather than preventing a single person from suffering a small risk of getting a minor headache. But if we keep stacking those small increments of harm to the one person, it is much less obvious that we should allow that one person to suffer constant, debilitating, life-long headaches in order to prevent however many thousands of people from suffering a minor headache. So perhaps we should not accept the NMD version of the avoid-greater-headaches principle, if we are inclined towards the sort of partially aggregative view which would tell us to prioritise the one person in this case, even when the options yielding the entire sequence of prospects are all available.[14]

[14] My discussion here ignores views which are entirely non-aggregative. There are two reasons for this. The first is that it seems to me that proponents of such views will not (and perhaps should not) be persuaded by any version of the Scale Argument for Deontic Longtermism. The second is that I find such views implausible.

This analogy stacks the decks against the NMD version of Future People Count a little too much, however. It's not minor headaches at stake for future people. It's one-in-ten-thousand risks of coming into existence with an atrocious life. We should not (if we take anti-aggregative intuitions seriously) treat these risks as being equivalent to the certainty of suffering a minor headache, even if the effects on ex ante wellbeing are similar. For suppose the probabilities of each future person coming into existence with an either atrocious or mediocre life were independent. In that case, choosing L would almost certainly prevent around 10^{20} future people from coming into existence with atrocious lives (though it would be highly uncertain who would be affected). It's clear enough that in such a case, one ought to choose L rather than S.

In fact, of course, the probabilities of the future people coming into existence with atrocious or mediocre lives are not independent of each other. There is a single state of nature, with a probability of one in ten thousand, in which all of the action occurs. My point is not that advocates of partial aggregation clearly should aggregate one-in-ten-thousand risks of grave harms in such a way that they eventually outweigh any number of certain grave harms. It is that, if we are going to accept a partially aggregative view, it is a mistake to treat all risks of harm which have the same ex ante magnitude in the same way; consequently, one cannot simply argue that, because we ought to prevent a single life-long debilitating headache rather than many thousands or millions of minor headaches, it immediately follows that the NMD version of *Future People Count* is false. More work, and more careful thought, is required than that.[15]

What about No Moral Dilemmas? I don't have anything particularly sophisticated to say in defence of this principle. But, in any choice situation, you've got to do something. If a moral theory tells you that everything you are able to do is wrong, that's not particularly helpful. Since we can't follow such a theory, I'm not sure that we should try.

The final premise involved in the argument is Obligation-Contraction Consistency. I have already explained the rough case for accepting it. But there is also a case for rejecting it. We have admitted that the relation $\succ_D$ might be intransitive. We might go further: we might think that it is *cyclic*, in the sense that there are options $O_1, O_2, \cdots, O_n$, such that $O_i \succ O_{i+1}$ and $O_n \succ O_1$. Suppose that is the case, and suppose further that in the option set $\{O_1, O_2, \cdots, O_n\}$, one of the options O_i is obligatory. In that case, Obligation-Contraction Consistency will imply that O_i is obligatory in the set $\{O_{i-1}, O_i\}$.[16] But *ex hypothesi*, that is not the case.

Again, a more concrete example might help to bring out this inconsistency. Consider Frick (2022)'s treatment of the Mere Addition Paradox in population ethics. Frick holds that, in a choice between bringing about a population A involving ten billion excellent lives, or instead bringing about A^+, in which the same ten billion people have *even better* lives, and additionally a large number of people exist with lives barely worth living, one ought to choose A^+. Between

[15] See Curran (2022) and Heikkinen (2022) on whether partial aggregation threatens the case for Longtermism.

[16] Or the set $\{O_n, O_1$, if $i = 1$.

A^+ and a population Z, in which the same people exist but with the wellbeing levels equalised and then raised slightly, so that each person has a life barely worth living (but better than those of the additional people in A^+), one ought to choose Z. But in a choice between A and Z, one ought to choose A. Frick additionally holds, quite plausible, that in a choice between all three, one ought to choose A because one has strong moral reason against bringing the additional people in A^+ into existence *if* those people could have been better off.

This plausible combination of judgements is inconsistent with Obligation-Contraction Consistency. For this principle implies that, since A is obligatory in the option set $\{A, A^+, Z\}$, it must also be obligatory in the option set $\{A, A^+\}$. *Ex hypothesi*, that is not the case.

Of course, such combinations of judgements might be wrong. There are powerful reasons to doubt deontic cyclicity which do not necessarily support Deontic Transitivity (at least, not to the same extent).[17] Less promisingly, one might think that even if there are deontic cycles, we can never find a set including all of the members of such a cycle with a single obligatory element. But given that one of the big philosophical payoffs of denying principles like Deontic Transitivity is precisely that we get to say what we want in cases like the Mere Addition Paradox, and given that we will often want to make cyclic judgements in these otherwise-paradoxical cases while maintaining that there is an option which ought to be chosen if all others in the cycle are available, there may not be much to be said for the half-way house of denying Deontic Transitivity while rejecting Deontic Cyclicity.

In full generality, then, and conditional on denying Deontic Transitivty, Obligation-Contraction Consistency looks dubious. Mainly for this reason, while the No Moral Dilemmas version of the Scale Argument for Deontic Longtermism looks compelling at first sight, I'm not sure that it is fully persuasive. Still, it's worth pointing out that even if Obligation-Contraction Consistency is not true in general, it might not lead us astray *in this case*. If this is the premise we reject, then we will still admit that one ought to do L rather than S if other, intermediate but unrealistic, choices are also available. It's hard to see how the presence of the intermediate choices could give us any *stronger* reason for doing L rather than S. My reaction is that, even Obligation-Contraction Consistency is dubious in general, it's quite unnatural to think that L is obligatory only in the presence of the intermediate options. I admit that this gut feeling is not much of an argument, but perhaps it is a gut feeling that you share.

[17] Gustafsson and Rabinowicz (2020) have provided a money pump for cyclic preferences which works even on agents who reason with foresight of their future choices. If choosing in accordance with morality is one way of choosing rationally, and if it is irrational to be money pumped, then in so far as this money pump succeeds, it shows that deontic cyclicity is false. Gustafsson (2022) argues that money-pump arguments can also be used to support transitivity. However, this argument is less secure than the money-pump argument against cyclicity.

4.2 The Agglomeration Approach

A second approach is due to Caspar Hare (2016). The basic idea is to take the intermediate prospects $X_1, X_2, \cdots$ as component parts of a single complex action. We then apply a principle to the effect that if we ought to perform all of the component parts of the action, then we ought to perform the complex action. Hence, we ought to choose L over S in *Short-Term vs Long-Term Intervention*.

That's the basic idea. To get the argument going, we will first need to slightly modify our principle that *Future People Count*:

> *Future People Count (Agglomeration version)* Suppose that prospects X and Y differ in only the following ways:
>
> (i) For some state of nature s_i with probability $\frac{1}{10{,}000}$, one present person has wellbeing level at most $w+1$ in X, rather than having wellbeing level w in Y, where $w \in [0, 99]$.
>
> (ii) In some state of nature s_j (which may not be equal to s_i) with probability $\frac{1}{10{,}000}$, at least one hundred million future people exist with wellbeing level -100 in prospect X. In prospect Y, these hundred million future people do not exist. Instead, one hundred million different future people exist with wellbeing level ϵ.
>
> Then, in a choice between X and Y, one ought, irrespective of one's previous choices, to choose Y.

We then need the following principle of

> *Agglomeration* For any sequence of actions $A_1, \cdots, A_n$, if, for each $1 \leq i \leq n$, you ought to do A_i, irrespective of whether or not you've done $A_1, \cdots A_{i-1}$, then you ought to do $A_1, \cdots, A_n$.

The argument proceeds as follows. Recall our prospects $X_0, X_1, \cdots, X_{10^{16}}$. We can imagine bringing about these prospects via a sequence of actions. Each one has a one in ten thousand chance of making a present person worse off, to the tune of one unit of wellbeing, in order to prevent a one in ten thousand chance (in the same state of nature each time) of one hundred million future people from having atrocious lives. Let's imagine that the effects of these actions are independent of each other, so that (for example) if one performs all but (any) one of the actions, then one hundred million future people will have a one in ten thousand chance of coming into existence with atrocious lives, while one lucky present person will have a one in ten thousand chance of having 1 unit of wellbeing, rather than ϵ units.

The Agglomeration version of *Future People Count* then implies that one ought to perform all of the actions in this sequence, irrespective of whether one has performed the previous actions. Agglomeration then implies that you ought to perform the entire sequence of actions. And that comes to the same thing as choosing L over S. With one exception. Once again, we are appealing to a case where there are a range of intermediate options technically available. One could

stop one action short of completing the entire sequence. Or, one could stop two actions short. And so on. In this way, all of the prospects $X_0, X_1, \cdots, X_{10^{16}}$ are available to the agent in this case, just as in the No Moral Dilemmas version of the Scale Argument for Deontic Longtermism.

Does this make a difference? Here's an argument that it doesn't. Imagine that all of the actions in the sequence can be carried out by pressing a series of buttons: one for each action. In that case, you want to press all of the buttons. And you want to do so regardless of which other buttons you have pressed: consequently, you can do it in whatever order you want, you just want the buttons pressed. Now imagine that in a straight choice between S and L, there are just two buttons: one for each choice. The only difference between these two cases is the number of buttons you need to press. Surely that couldn't make any significant moral difference.

It remains to be seen whether the Agglomeration version of *Future People Count*, and the Agglomeration principle, should be accepted. The previously mentioned objections to Future People Count in the No Moral Dilemmas case also apply here. In particular, it *might* be that on the best partially aggregative moral theories, applied to cases of risk, *Future People Count* is false because it effectively requires us to prevent tiny chances of large harms to the relatively many, rather than preventing large and certain harms to the relatively few.[18] But then again, the best partially aggregative moral theories might not say this. Given the choice to either prevent a single premature death for sure, or instead prevent a one in ten thousand risk of a genocide of hundreds of millions, it is hardly obvious that one ought to prevent the single premature death.

As for Agglomeration, it's hard to see on what grounds one might reject this principle. Consider the two claims that

(i) one ought to press all ten quadrillion buttons individually, irrespective of which other buttons one may or may not have pressed, but

(ii) it is not the case that one ought to press all ten quadrillion buttons.

Taken together, these two claims are of dubious coherence.

5 Conclusion

Longtermists claim that there could be a lot of future people. And they argue that *because* the future might contain so many people, the effects of our actions on the long run (including the indirect ones) are more non-instrumentally important, morally speaking, than the effects of our actions on the short run. I have tried to flesh out this Scale Argument for Longtermism in some more detail. We first applied the Scale Argument to the case of Axiological Longtermism. We

[18] Note that if we carry out L, we prevent potential harms to future people by removing the possibility of their existence. Plausibly, this does not make much difference to our moral reasons. It is widely believed that we have strong moral reasons to avoid creating people with miserable lives; see for instance Frick (2014), Roberts (2011) or McMahan (2009).

found that Axiological Longtermism follows from the rather weak claim that *Future People Count*, in the sense that averting small chances of grave harms to huge numbers of future people is better than averting same-sized chances of relatively trivial harms to individual present people. At least, this is so on the (correct) assumption that the betterness relation is transitive.

We then turned to the deontic case. We first examined some reasons to doubt that we should accept a deontic version of transitivity. We then attempted to modify the Scale Argument so as to do without a transitivity assumption. The "No Moral Dilemmas Approach" replaced transitivity with the principle that there is always at least one permissible member of an option set, and slightly modified the claim that *Future People Count*. With this argument, we were able to find that one ought to perform a long-term intervention rather than a short-term one *provided* a great many intermediate options were available. In practice, they are not. Our conclusion could be extended to the pairwise case where there are no intermediate options using the principle of Obligation-Contraction Consistency. However, we saw some reasons to be sceptical of this principle. Overall, then, the No Moral Dilemmas version of the Scale Argument for Deontic Longtermism was found to have some force, but to fall short of being fully persuasive.

We next looked at another way of extending the Scale Argument to the deontic case. The "Agglomeration" approach replaced transitivity with the Agglomeration principle, which states that if one ought to perform each member of a sequence of actions, irrespective of which other members of the sequence one may or may not have performed, the none ought to perform the entire sequence of actions. I argued that this Agglomeration principle should be accepted. This version of the Scale Argument then led to the conclusion that one ought to choose a long-term intervention over a short-term one if one can perform the long-term intervention by pressing ten quadrillion separate buttons. Again, this is not the case in practice. However, I argued that since it does not matter whether we perform an action by pressing ten quadrillion buttons or by pressing one button, we should nevertheless conclude that one ought to choose the long-term intervention L over the short-term intervention S in a pairwise choice between the two, provided we accept the relevant version of *Future People Count*.

Should we accept *Future People Count*? I tentatively think that we should, for all of the versions presented in this paper. Perhaps it makes sense to give significant priority to present people over possible people who may or may not exist. But I don't think it makes sense for the level of priority to be arbitrarily large. The best way to reject *Future People Count*, in the forms needed to get the deontic Scale Arguments going, is probably to appeal to a theory of partial aggregation. But it is, as of yet, unclear what the best such theory will say about cases of risk. I tend to think that the best partially aggregative moral theories will tell us that small risks of grave harms to the many can compete with large risks of grave harms to the few. If I'm wrong about that, then *Future People Count* may be on shakier ground than I imagine.

Let me mention a final worry one might have with the claim that L is better than S (or that one ought to choose L), namely that this claim is so implausible

that it counts as a *reductio*. Perhaps there is something to be said for this way of looking at things when we imagine that L effects an extinction-mitigating change. I think there is less to be said for looking at things in this way when we keep in mind that L effects a suffering-reducing change.

References

Arrhenius, G. 2022. Population paradoxes without transitivity. In *The Oxford Handbook of Population Ethics*, eds. G. Arrhenius, K. Btykvist, T. Campbell, and E. Finneron-Burns, Chapter 8. New York: Oxford University Press.

Boonin, D. 2014. *The non-identity problem and the ethics of future people*. New York: Oxford University Press.

Bostrom, N. 2003. Astronomical waste: The opportunity cost of delayed technological development. *Utilitas 15*(3): 308–314.

Broome, J. 2004. *Weighing Lives*. Oxford: Oxford University Press.

Broome, J. 2005. Should we value population? *Journal of Political Philosophy 13*(4): 399–413.

Chappell, R. Y. 2017. Rethinking the asymmetry. *Canadian Journal of Philosophy 47*(2-3): 167–177.

Curran, E. J. 2022. Longtermism, aggregation, and catastrophic risk. GPI Working Paper No. 18-2022.

Finneron-Burns, E. 2017. What's wrong with human extinction? *Canadian Journal of Philosophy 47*(2–3): 327–343.

Frick, J. 2014. *'Making People Happy, Not Making Happy People': A Defense of the Asymmetry Intuition in Population Ethics*. Ph.D. thesis, Harvard University.

Frick, J. 2017. On the survival of humanity. *Canadian Journal of Philosophy 47*(2-3): 344–367.

Frick, J. 2022. Context-dependent betterness and the mere addition paradox. In *Ethics and Existence: the Legacy of Derek Parfit*, eds. J. McMahan, T. Campbell, J. Goodrich, and K. Ramakrishnan, Chapter 9, 232–263. Oxford: Oxford University Press.

Frick, J. and H. Lederman. forthcoming. Response to "the case for strong longtermism". In *Essays On Longtermism*, eds. J. Barrett, D. Thorstad, and H. Greaves. Oxford University Press.

GiveWell 2022. Givewell's cost-effectiveness analyses. Accessed 2nd February 2023.

Greaves, H. 2016. Cluelessness. *Proceedings of the Aristotelian Society 116*(3): 311–339.

Greaves, H. and W. MacAskill. 2021. The case for strong longtermism. GPI Working Paper No. 5–2021.

Gustafsson, J. E. 2022. *Money-Pump Arguments*. Cambridge: Cambridge University Press.

Gustafsson, J. E. and W. Rabinowicz. 2020. A simpler, more compelling money pump with foresight. *The Journal of Philosophy 117*(10): 578–589.

Hare, C. 2016. Should we wish well to all? *The Philosophical Review 125*(4): 451–472.

Heikkinen, K. 2022. Strong longtermism and the challenge from anti-aggregative moral views. GPI Working Paper No. 5-2022.

Kaneda, T. and C. Haub. 2022. How many people have ever lived on earth? Population Reference Bureau.

Lazar, S. and C. Lee-Stronach. 2019. Axiological absolutism and risk. *Noûs 53*(1): 97–113.

Lenman, J. 2000. Consequentialism and cluelessness. *Philosophy & Public Affairs 29*(4): 342–370.

MacAskill, W. 2022. *What We Owe The Future*. New York: Basic Books.

McMahan, J. 2009. Asymmetries in the morality of causing people to exist. In *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, eds. M. A. Roberts and D. T. Wasserman, Chapter 3, 49–68. Dordrecht: Springer.

Narveson, J. 1967. Utilitarianism and new generations. *Mind 76*(301): 62–72.

Newberry, T. 2021. How many lives does the future hold? GPI Technical Report No. T2-2021.

Ord, T. 2020. *The Precipice: Existential Risk and the Future of Humanity*. London: Bloomsbury.

Roberts, M. A. 2011. The Asymmetry: A Solution. *Theoria 77*(4): 333–367.

Thorstad, D. 2022. Existential risk pessimism and the time of perils. GPI Working Paper No. 1-2022.